Video: Protein-Protein Interactions
Credit: Mauro Martino

Network Science

dataScience UDD

NICO NORTHWESTERN INSTITUTE ON COMPLEX SYSTEMS

Cristian Candia-Castro Vallejos, Ph.D.

cristiancandia@udd.cl

Director Magister en Data Science UDD

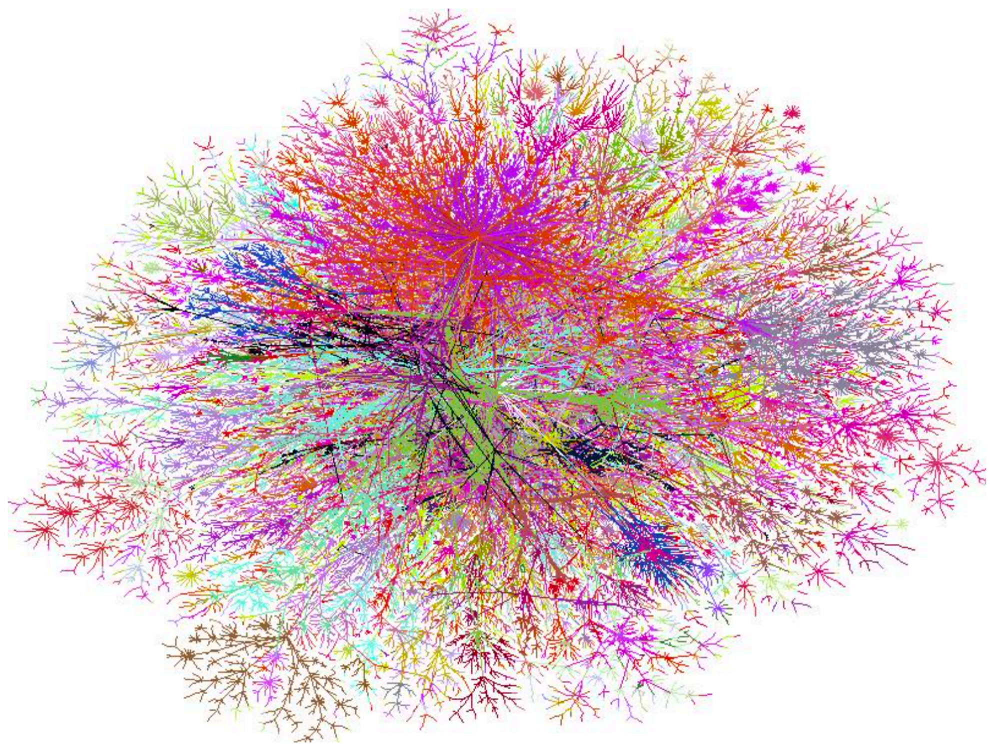
Profesor Investigador, Facultad de Ingeniería, UDD

External Faculty Northwestern Institute on Complex Systems,
Kellogg School of Management, Northwestern University

Centralidad

Estas diapositivas se basan en la presentación original del
Prof. Albert-László Barabási, de Northeastern University, con autorización.
El contenido ha sido traducido para su uso en este curso.

Las redes son una representación matemática del mundo real



H. Burch and B. Cheswick

Simplificar la complejidad

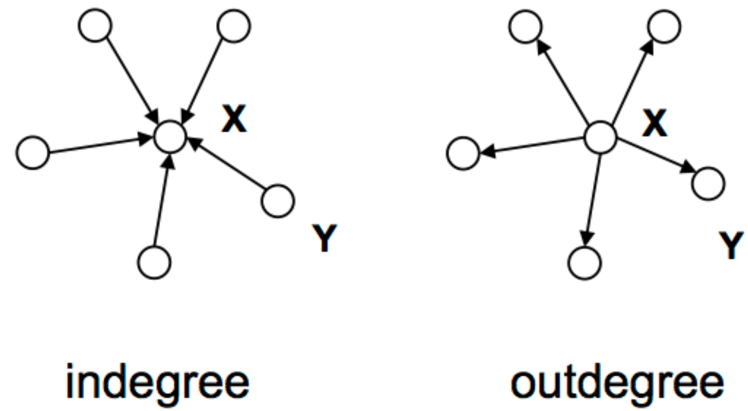
- **Patrones generales:** centralidad de grado, longitud de camino…
- **Patrones locales:** coeficiente de agrupamiento, transitividad…
- **Posición en la red:** comunidades, centralidad, influencia…

Medidas de centralidad

- **Posición en la red:** centralidad de grado, intermediación, cercanía
- **Influencia:** centralidad eigenvector, Katz, PageRank
- **Mecanismos:** in-degree, out-degree, flujo

Centralidad

- **Centralidad es una medida local de la red**
- Existen distintas medidas de centralidad, que se deben elegir en función de la mejor representación del sistema:
  - cuantas personas conoces?
  - con quien compartes -envias- tus recursos?
  - que tan popular eres?



Going the Wrong Way on a One-Way Street: Centrality in Physics and Biology*

Linton C. Freeman, lin@aris.ss.uci.edu
University of California, Irvine

Abstract

When ideas and tools move from one field to another, the movement is generally from the natural to the social sciences. In recent years, however, there has been a major movement in the opposite direction. The idea of centrality and the tools for its measurement were originally developed in the social science field of social network analysis. But currently the concept and tools of centrality are being used widely in physics and biology. This paper examines how and why that—wrong way—movement developed, its extent and its consequences for the fields involved.

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOLUME 22, NUMBER 6 NOVEMBER, 1950

Communication Patterns in Task-Oriented Groups

ALEX BAVELAS
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received August 8, 1950)

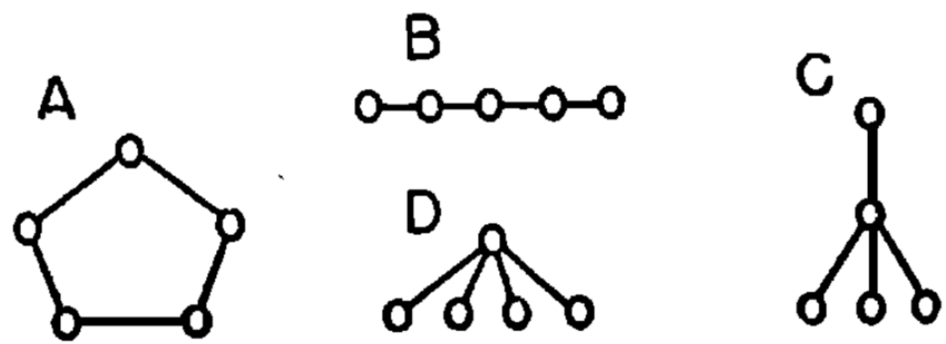


FIG. 1. Some illustrative communication patterns among five individuals.

Social Networks, 1 (1978/79) 215–239

Centrality in Social Networks Conceptual Clarification

Linton C. Freeman

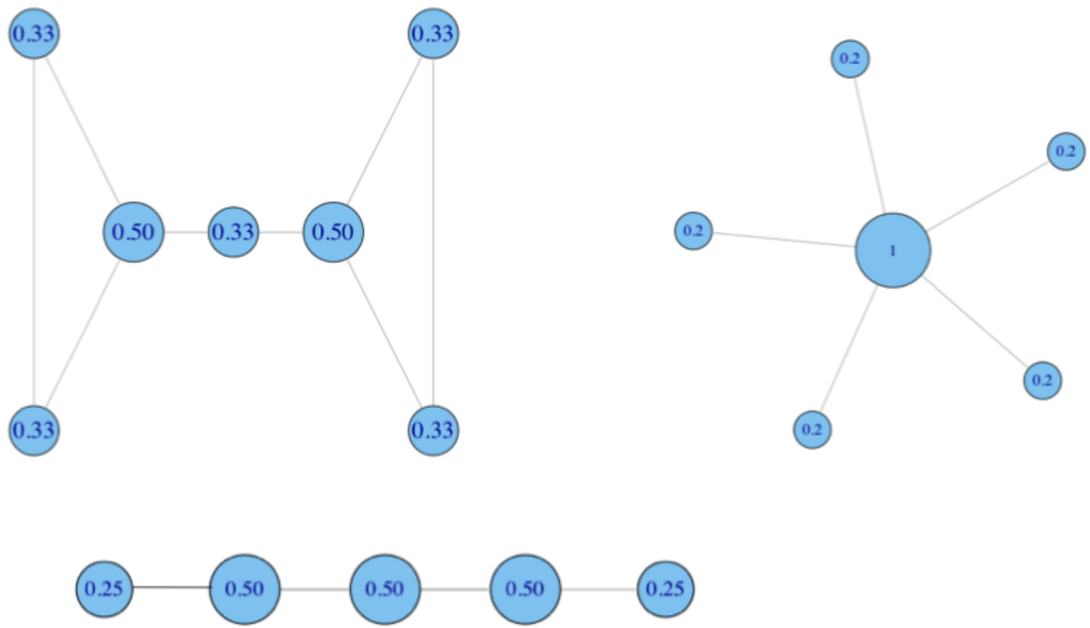
*Lehigh University**

The intuitive background for measures of structural centrality in social networks is reviewed and existing measures are evaluated in terms of their consistency with intuitions and their interpretability.

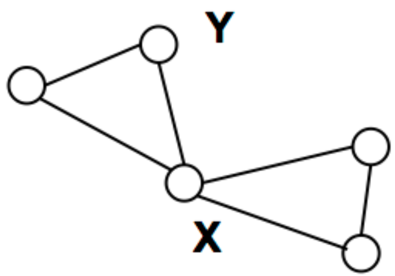
Three distinct intuitive conceptions of centrality are uncovered and existing measures are refined to embody these conceptions. Three measures are developed for each concept, one absolute and one relative measure of the centrality of positions in a network, and one reflecting the degree of centralization of the entire network. The implications of these measures for the experimental study of small groups is examined.

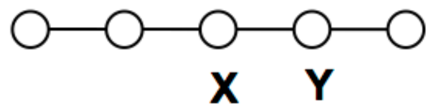
Centralidad de grado

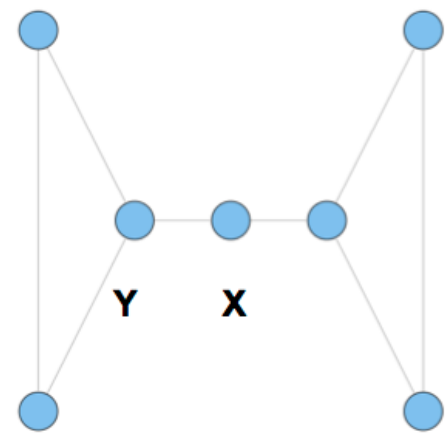
$$C^D(i) = \frac{k_i}{N-1}$$



Quien es mas importante, X o Y?







- Intermediacion entre grupos
- Probabilidad de que la información te llegue

Centralidad de intermediación

$$\tilde{C}^B(i) = \sum_{j<k} \frac{d_{jk}(i)}{d_{jk}}$$

d_{jk} Número de caminos mas cortos entre j y k

$d_{jk}(i)$ Número de caminos mas cortos entre j y k que pasan a traves de i

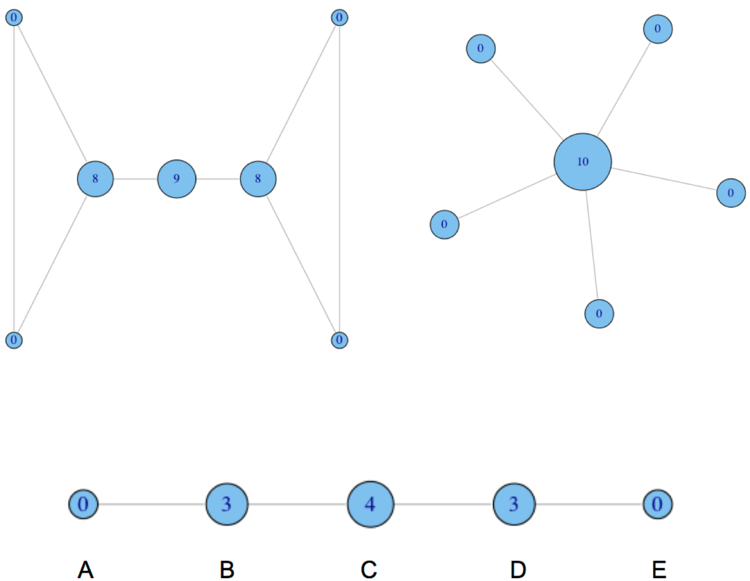
Centralidad de intermediación

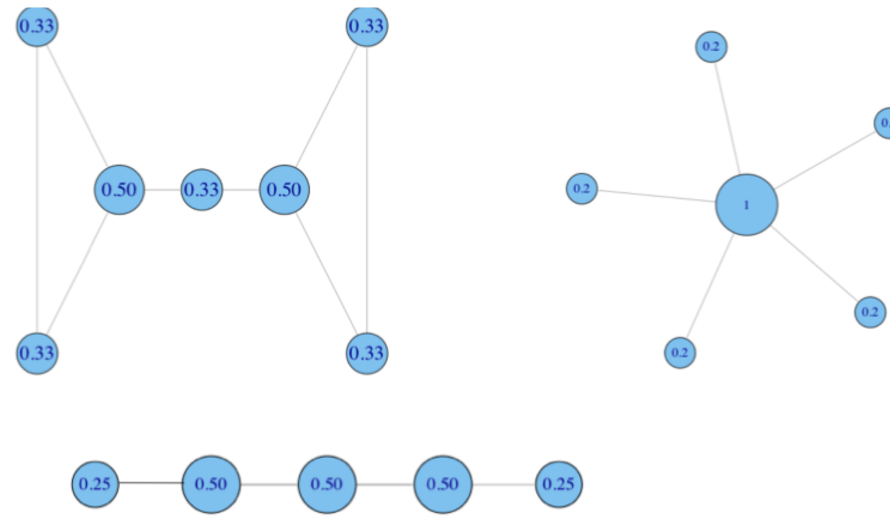
- Normalizacion:

$$C^B(i) = \frac{\tilde{C}^B}{(N-1)(N-2)/2}$$

- Denominador: numero de pares de vertices excluyendo *i*

Intermediacion vs. Grado





Centralidad de cercanía

$$\tilde{C}^{C}(i) = \left[\sum_{j=1}^{N} d(i,j)\right]^{-1} \qquad C^{C}(i) = \frac{\tilde{C}^{C}(i)}{N-1}$$

Centralidad eigenvector

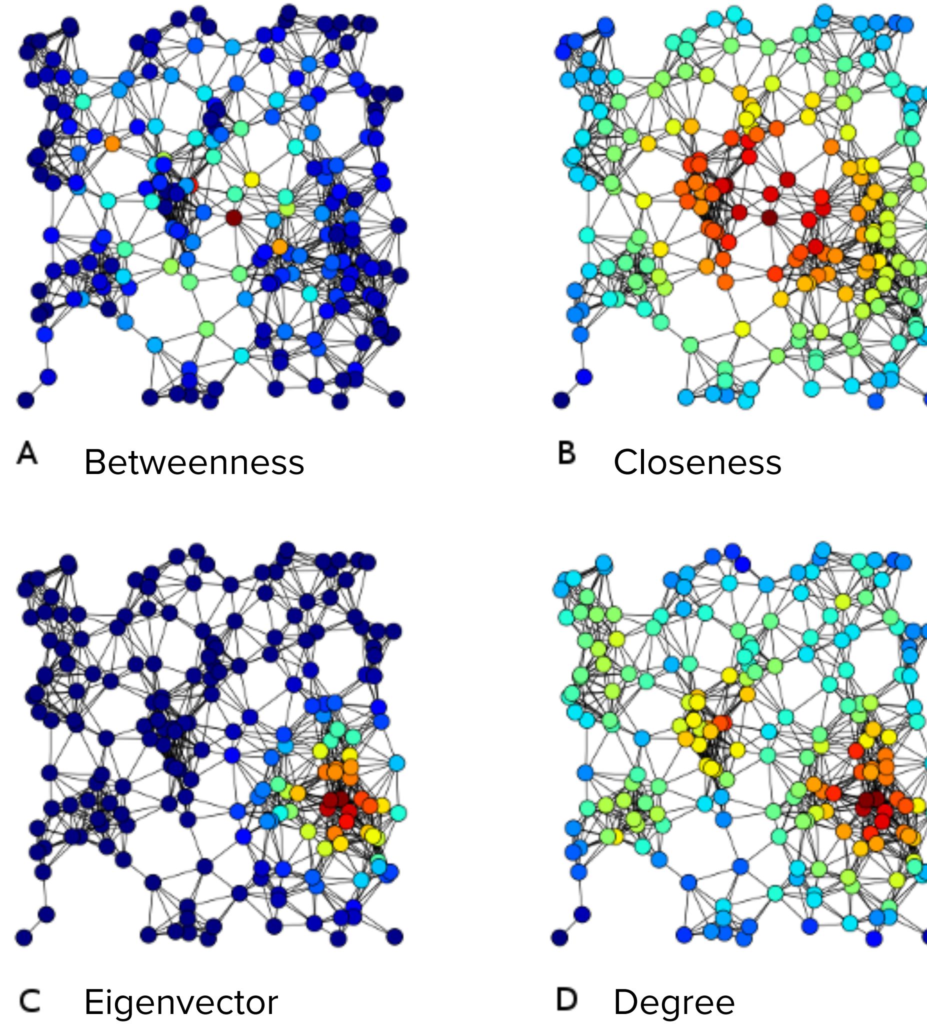
Nota: un nodo es importunate si esta conectado a otros nodos importantes

$$x_i = \frac{1}{\lambda} \sum_{j \in \Lambda(i)} x_j$$

$$x_i = \frac{1}{\lambda} \sum_{j \in G} a_{ij} x_j$$

$$AX = \lambda X$$

- Que hace a un nodo ser **importante**?
    - subjective ➜ muchas definiciones y sujeto al contexto en que se estudia

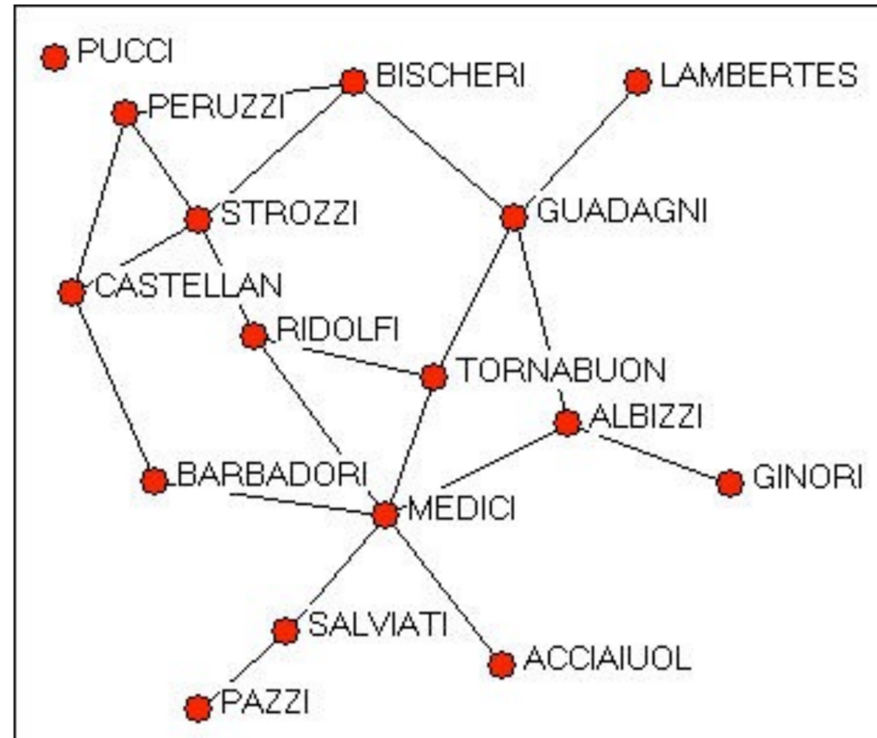


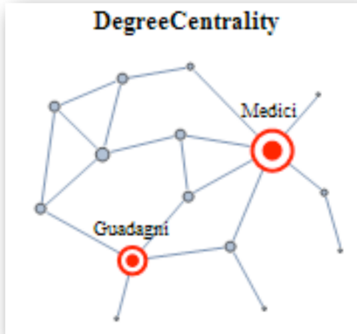
Robust Action and the Rise of the Medici, 1400–1434[1]

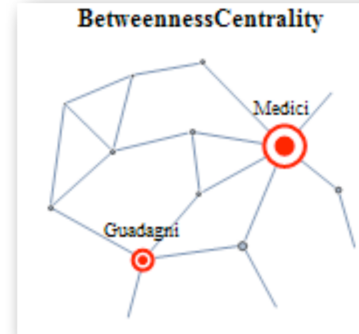
John F. Padgett and Christopher K. Ansell
University of Chicago

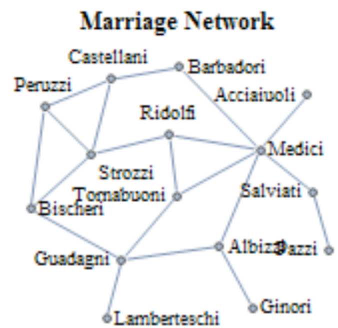
We analyze the centralization of political parties and elite networks that underlay the birth of the Renaissance state in Florence. Class revolt and fiscal crisis were the ultimate causes of elite consolidation, but Medicean political control was produced by means of network disjunctures within the elite, which the Medici alone spanned. Cosimo de' Medici's multivocal identity as sphinx harnessed the power available in these network holes and resolved the contradiction between judge and boss inherent in all organizations. Methodologically, we argue that to understand state formation one must penetrate beneath the veneer of formal institutions, groups, and goals down to the relational substrata of peoples' actual lives. Ambiguity and heterogeneity, not planning and self-interest, are the raw materials of which powerful states and persons are constructed.

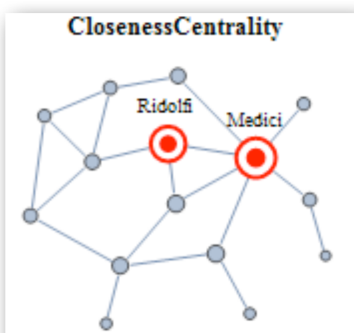
Familias florentinas

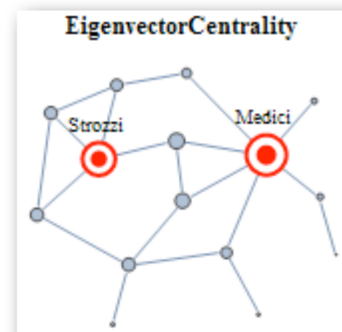



DegreeCentrality
Medici
Guadagni

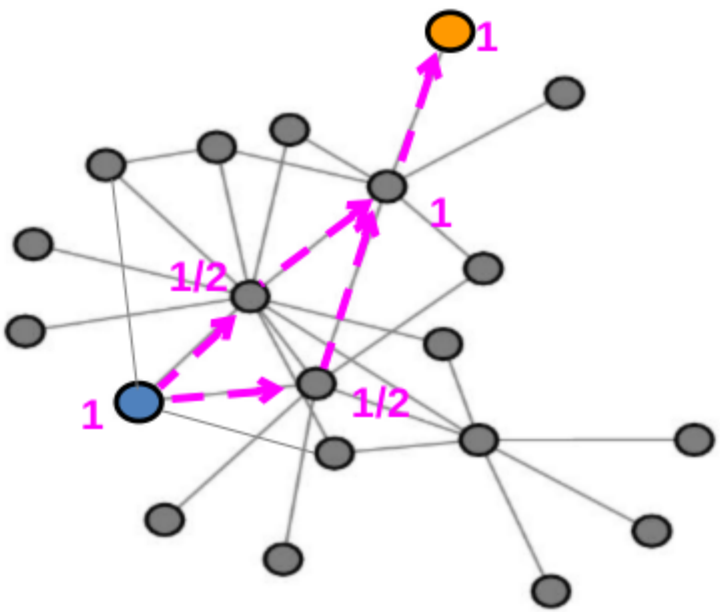

BetweennessCentrality
Medici
Guadagni


Marriage Network
Castellani
Barbadori
Peruzzi
Acciaiuoli
Ridolfi
Medici
Strozzi
Tornabuoni
Salviati
Bischeri
Guadagni
Albizzi
Pazzi
Ginori
Lamberteschi


ClosenessCentrality
Ridolfi
Medici


EigenvectorCentrality
Strozzi
Medici

De vuelta a centralidad de intermediación



- **Centralidad de carga (Load)**: el numero de paquetes que un nodo debe manejar durante el transporte basado en el camino más corto desde todos los nodos hacia todos los demás
- A que medida de centralidad se parece?

Centralidad en redes con peso

Por qué el peso es importante

The Strength of Weak Ties[1]

Mark S. Granovetter
Johns Hopkins University

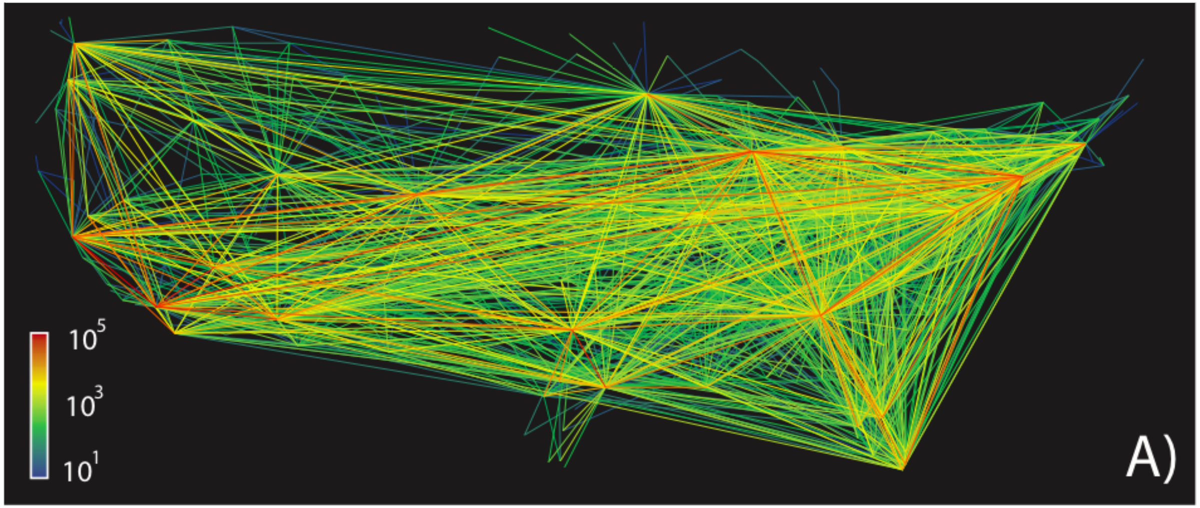
Analysis of social networks is suggested as a tool for linking micro and macro levels of sociological theory. The procedure is illustrated by elaboration of the macro implications of one aspect of small-scale interaction: the strength of dyadic ties. It is argued that the degree of overlap of two individuals' friendship networks varies directly with the strength of their tie to one another. The impact of this principle on diffusion of influence and information, mobility opportunity, and community organization is explored. Stress is laid on the cohesive power of weak ties. Most network models deal, implicitly, with strong ties, thus confining their applicability to small, well-defined groups. Emphasis on weak ties lends itself to discussion of relations *between* groups and to analysis of segments of social structure not easily defined in terms of primary groups.

Ejemplo 1: tráfico aéreo

Nodos: aeropuertos

Enlaces: vuelos directos

Peso: Numero de asientos

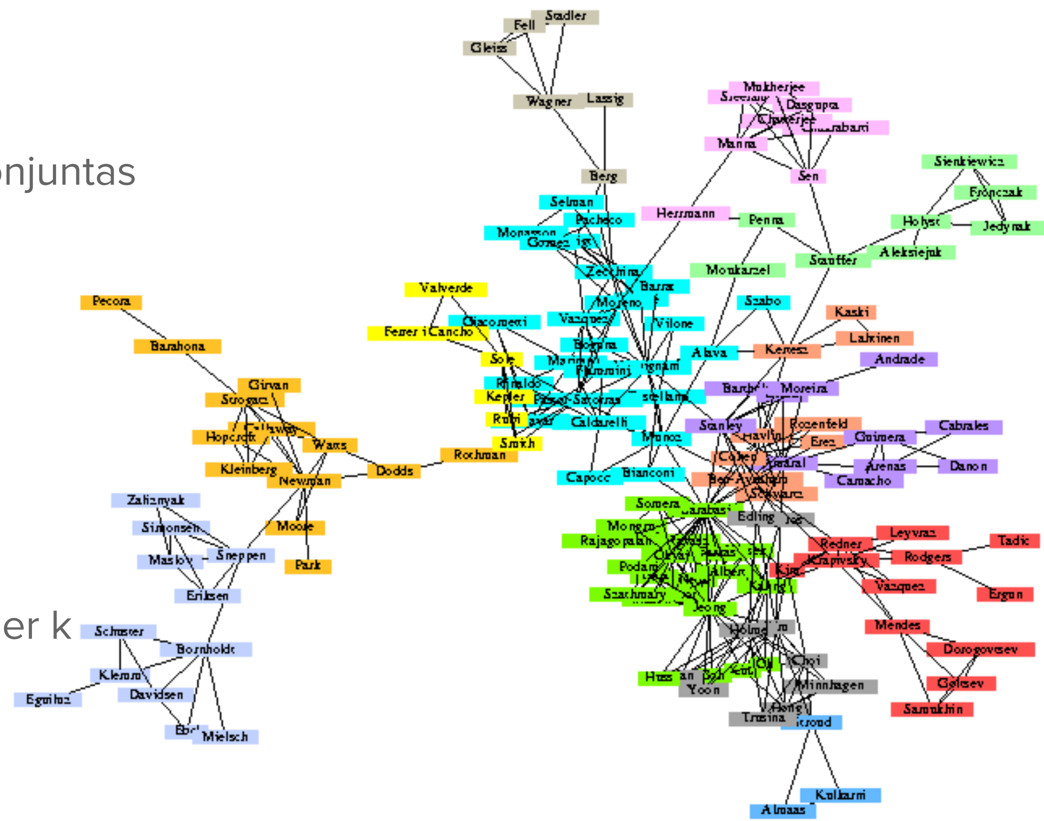


Ejemplo 2: Colaboraciones cientificas

- Nodos: científicos
- Enlces: publicaciones conjuntas
- Peso: numero de publicaciones conjuntas

$$w_{ij} = \Sigma_k \frac{\delta_i^k \delta_j^k}{n_k - 1}$$

- *k*: paper
- n_k numero de autores
- $\delta_{ik}= 1$ si el autor i contribuyo al paper k



The architecture of complex weighted networks

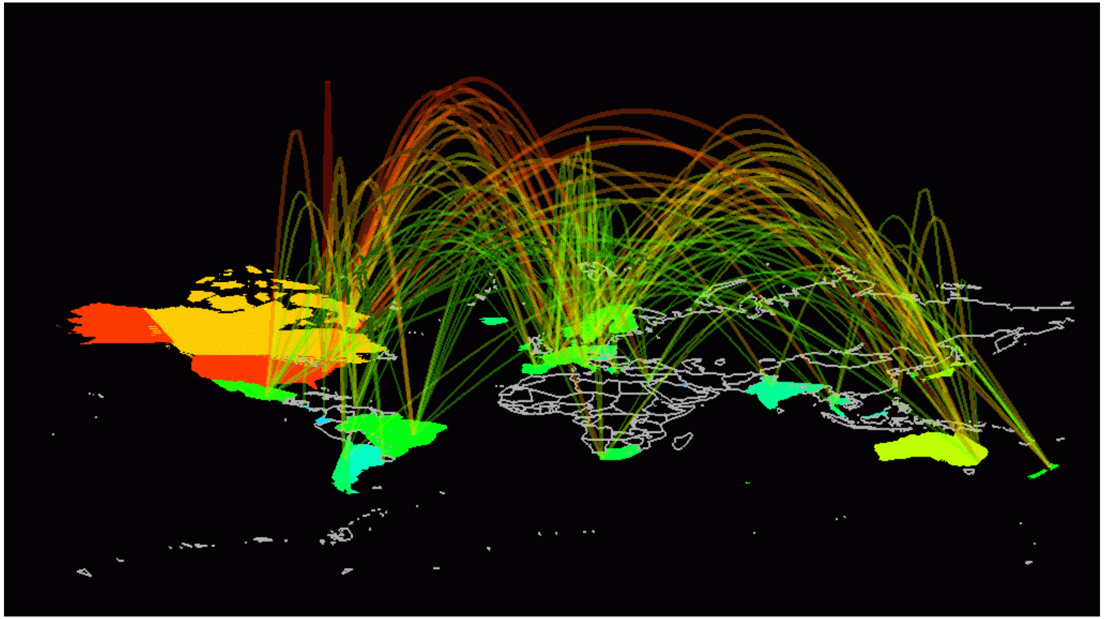
A. Barrat*, M. Barthélemy†, R. Pastor-Satorras‡, and A. Vespignani*§

*Laboratoire de Physique Théorique (Unité Mixte de Recherche du Centre National de la Recherche Scientifique 8627), Bâtiment 210, Université de Paris-Sud, 91405 Orsay Cedex, France; †Commissariat à l'Energie Atomique–Département de Physique Théorique et Appliquée, 91191 Bruyères-Le-Chatel, France; and ‡Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain

Communicated by Giorgio Parisi, University of Rome, Rome, Italy, January 8, 2004 (received for review October 29, 2003)

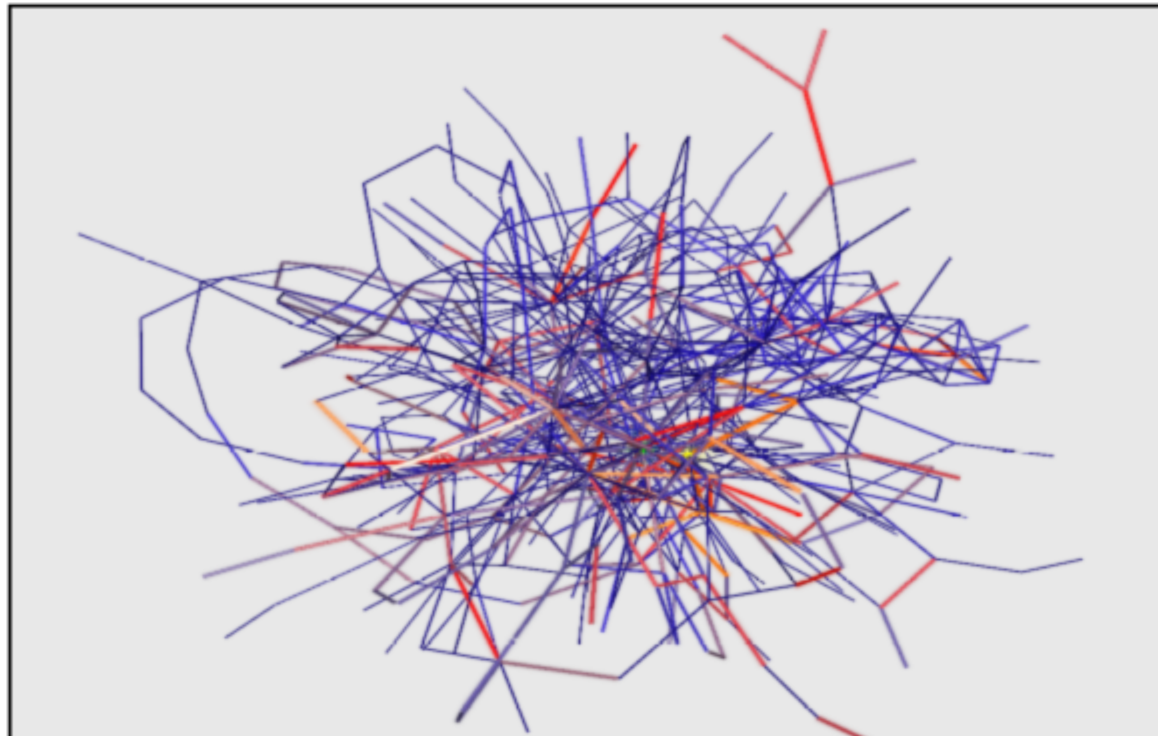
Ejemplo 3: Internet

- Nodos: routers
- Enlases: lineas físicas
- Peso: trafico



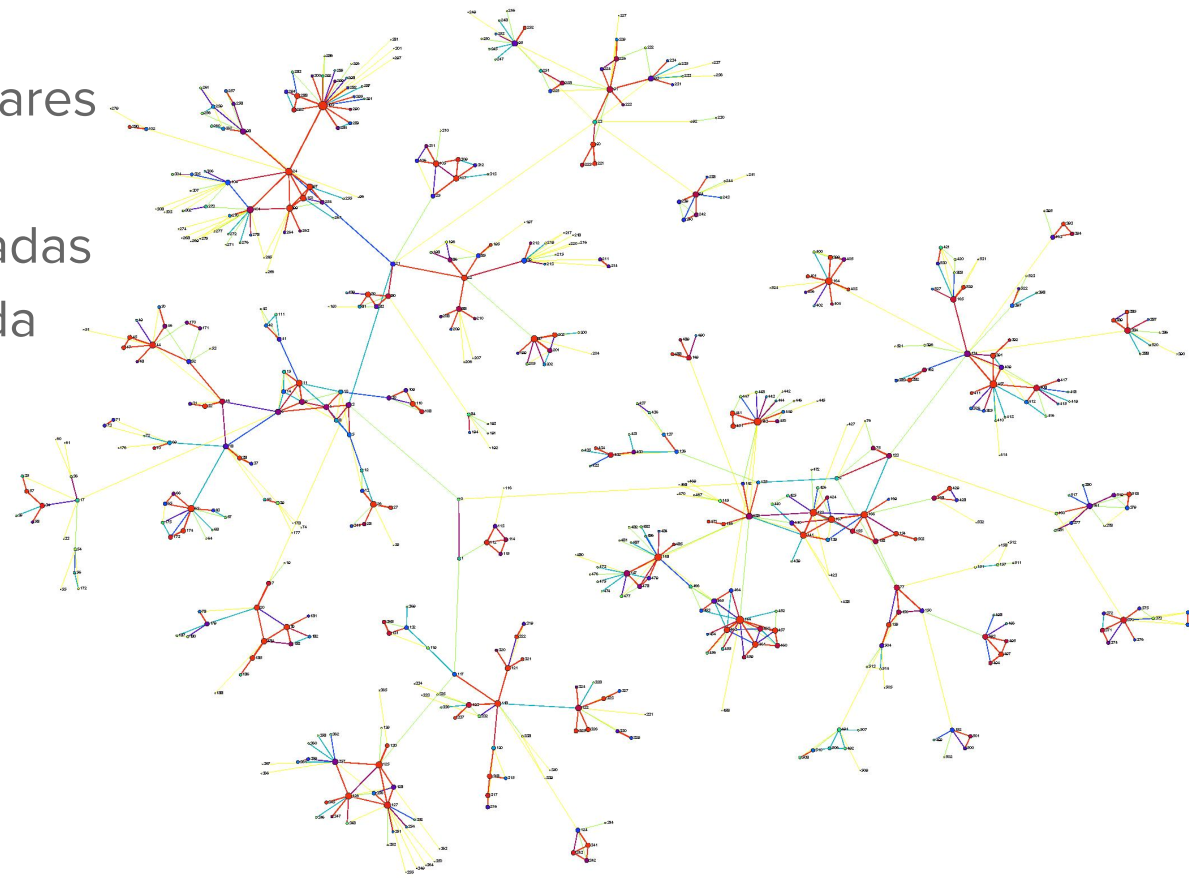
Ejemplo 4: Red metabólica

- Nodos: metabolitos
- Enlaces: reacciones
- Peso: flujo de reacción



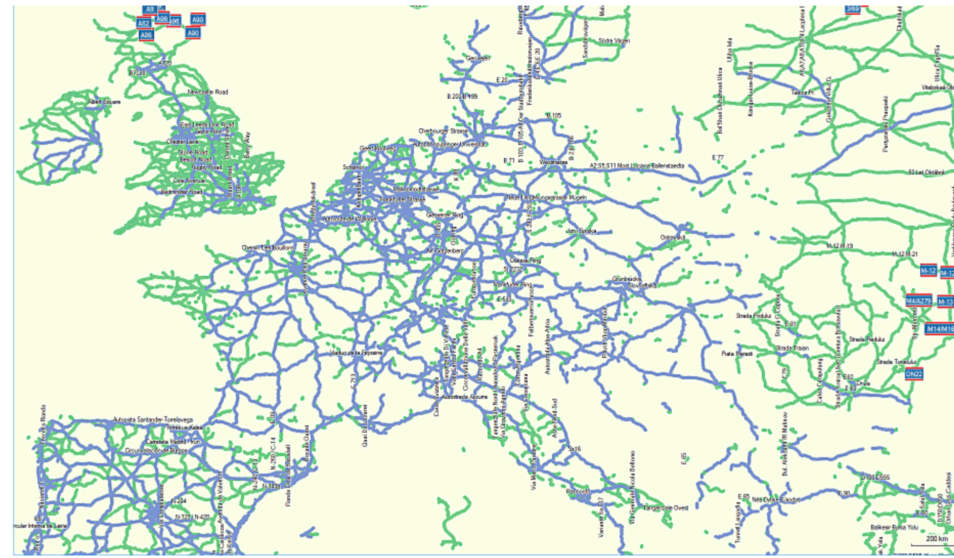
Ejemplo 5: Red de llamadas celulares

- Nodes: telefonos celulares
- Enlaces: llamadas
- Peso: numero de llamadas o duration de la llamada



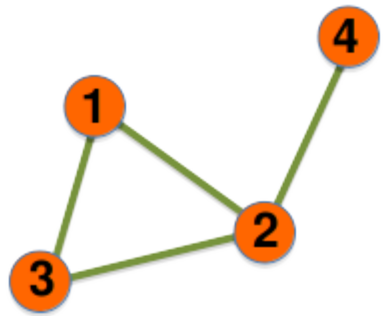
Qué significan los pesos?

- Es una decision critica de la moderación de la red
- Pueden significar:
  - distancia
  - costos
- O también:
  - mas conexiones
  - mayor correlación



En un caso, significa que el peso altos correspondie a longitudes de ruta largas, en el otro a longitudes de ruta cortas.

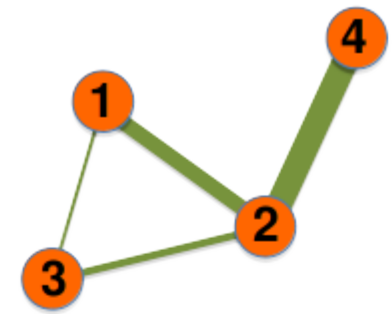
Grado con peso



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad \langle k \rangle = \frac{2L}{N}$$

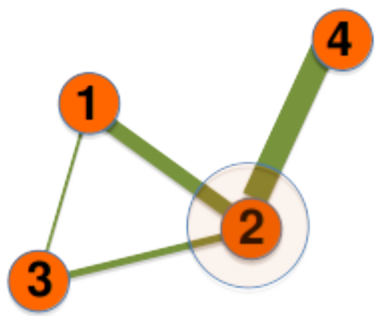


$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad \langle k \rangle = \frac{2L}{N}$$

a_{ij} y w_{ij}



Adjacency Matrix (A_{ij})

$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Weights matrix (W_{ij})

$$W_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

Node Strength (*weighted degree*) s:

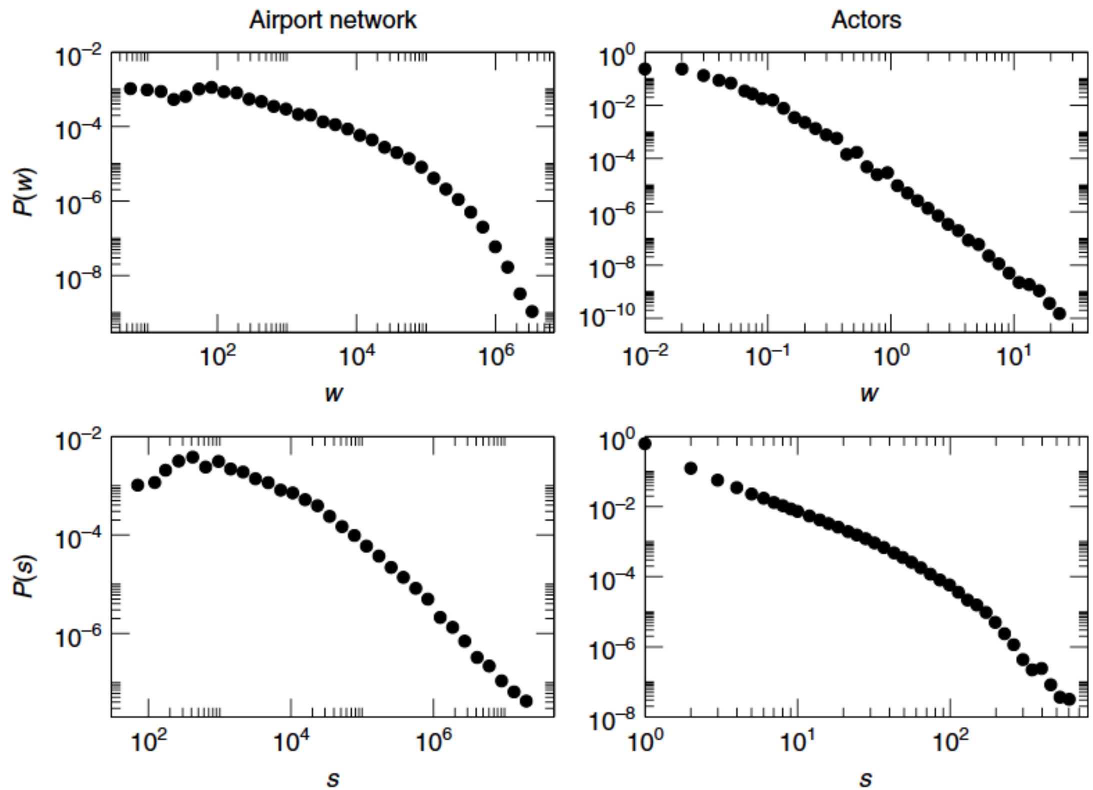
$$s_i = \sum_{j=1}^{N} a_{ij} w_{ij} = \sum_{j=1}^{N} w_{ij}$$

$$s_2 = \sum_{j=1}^{N} w_{2j} = w_{21} + w_{23} + w_{24}$$

P(s) y P(w) a menudo siguen una distribución de cola ancha

Strength distribution *P(s)*:
probability that a randomly chosen node has strength *s*

Weight distribution ***P(w)***:
probability that a randomly chosen link has weight ***w***



Importancia de la conectividad en redes Barabasi-Albert

Para generar redes libres de escala, Laszlo Barabási y Réka Albert sugirieron el algoritmo

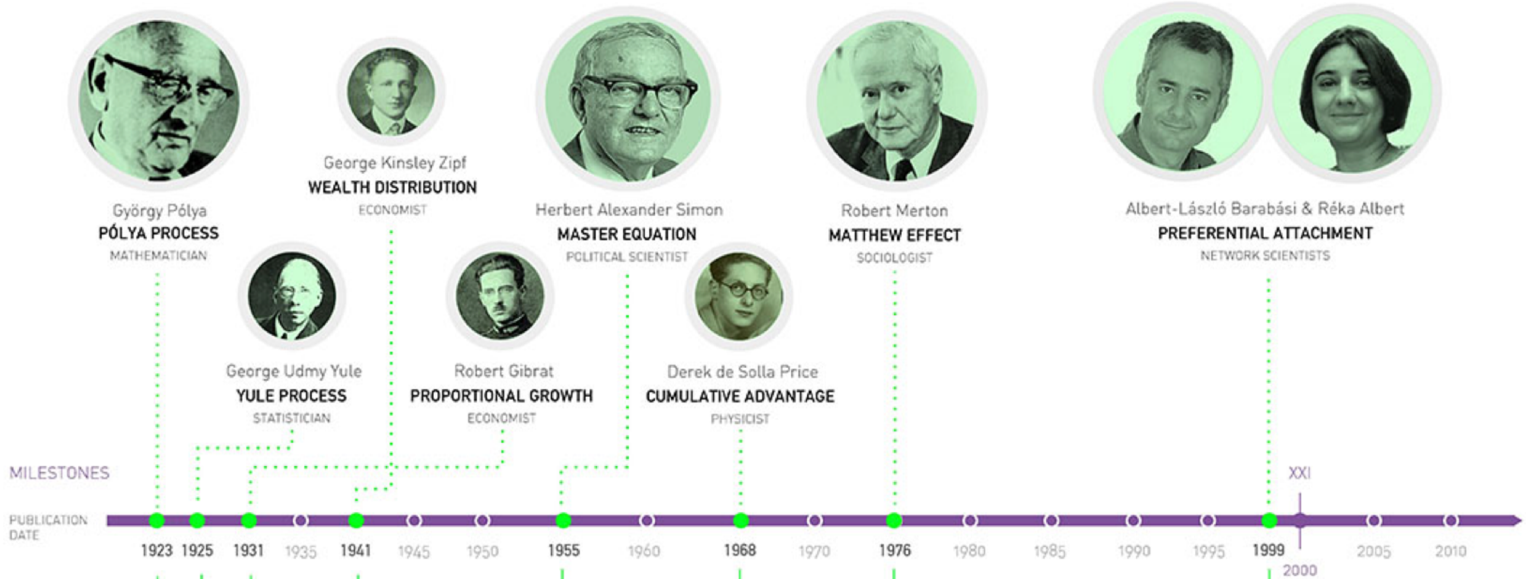
Barabási-Albert model of Growth and Preferential Attachment

Growth and Preferential Attachment

Crecimiento significa que tendremos una red en la que el número de nodos crecerá iterativamente en el tiempo hasta que alcancemos el tamaño de población deseado.

Los nuevos nodos se unirán a los nodos existentes, pero lo harán preferentemente a los más conectados.

Preferential Attachment



György Pólya (1887-1985)
Preferential attachment made its first appearance in 1923 in the celebrated urn model of the Hungarian mathematician György Pólya [2]. Hence, in mathematics preferential attachment is often called a **Pólya process**.

George Udmy Yule (1871-1951)
used preferential attachment to explain the power-law distribution of the number of species per genus of flowering plants [3]. Hence, in statistics preferential attachment is often called a **Yule process**.

Robert Gibrat (1904-1980)
proposed that the size and the growth rate of a firm are independent. Hence, larger firms grow faster [4]. Called **proportional growth**, this is a form of preferential attachment.

George Kinsley Zipf (1902-1950)
used preferential attachment to explain the fat tailed distribution of wealth in the society [5].

Herbert Alexander Simon (1916-2001)
used preferential attachment to explain the fat-tailed nature of the distributions describing city sizes, word frequencies, or the number of papers published by scientists [6].

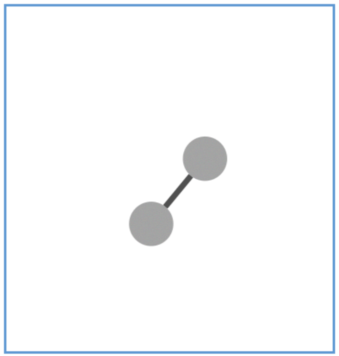
Derek de Solla Price (1922-1983)
used preferential attachment to explain the citation statistics of scientific publications, calling it **cumulative advantage** [7].

Robert Merton (1910-2003)
In sociology preferential attachment is often called the **Matthew effect**, named by Merton [8] after a passage in the Gospel of Matthew.

Barabási (1967) & **Albert** (1972)
introduce the term **preferential attachment** to explain the origin of scale-free networks [1].

Growth and Preferential Attachment

Step 0



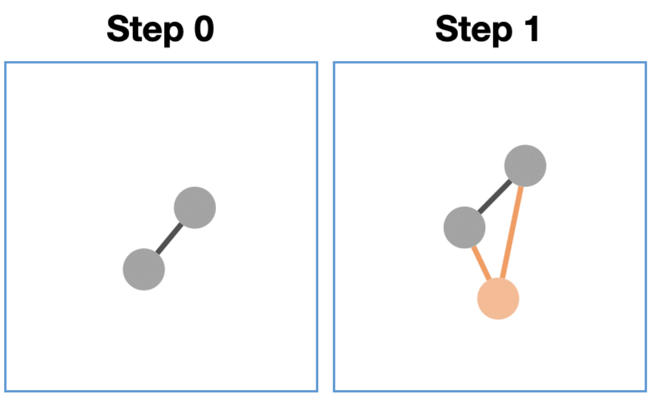
El algoritmo consta de los siguientes pasos:

- Comenzando con m0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las N-m0 iteraciones, agregue un nodo nuevo que se conecte con m (< m0) enlaces a nodos preexistentes con probabilidad:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

la probabilidad de que un enlace del nuevo nodo se conecte al nodo i depende del grado ki como

Growth and Preferential Attachment



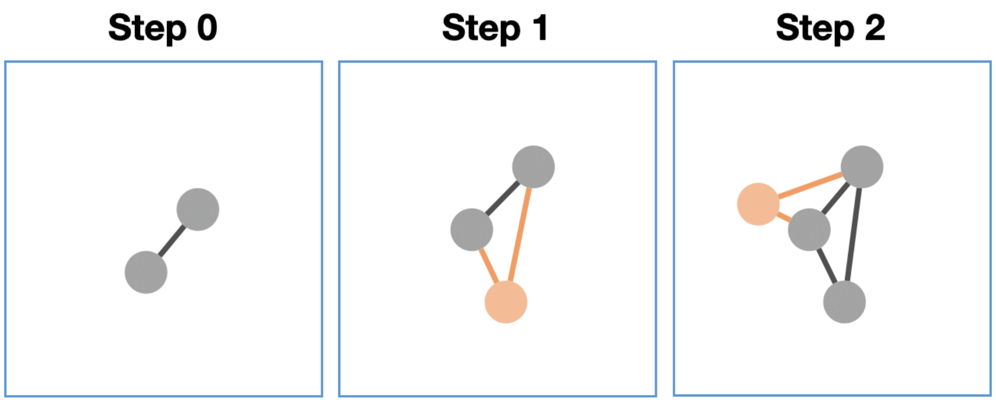
El algoritmo consta de los siguientes pasos:

- Comenzando con m0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las N-m0 iteraciones, agregue un nodo nuevo que se conecte con m (< m0) enlaces a nodos preexistentes con probabilidad:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

la probabilidad de que un enlace del nuevo nodo se conecte al nodo i depende del grado ki como

Growth and Preferential Attachment



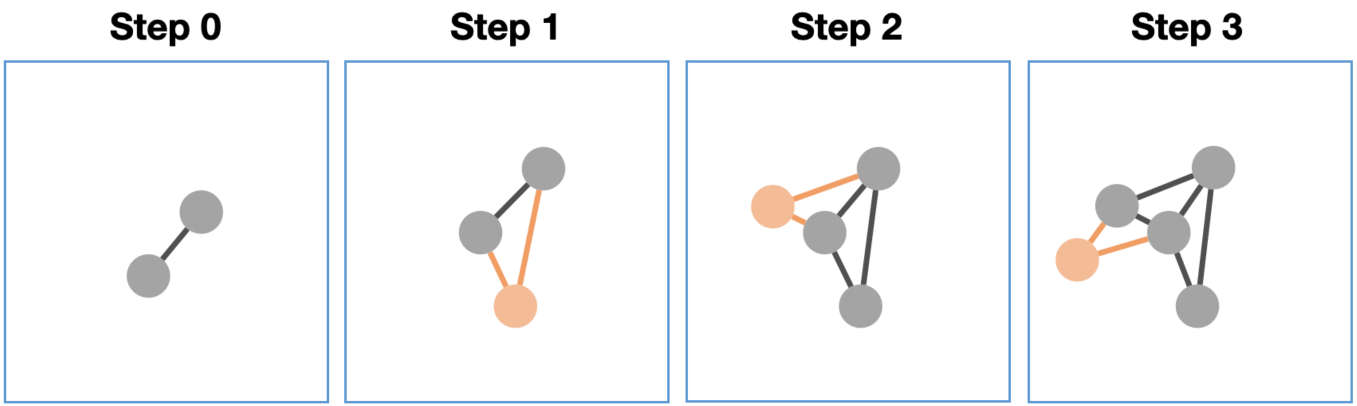
El algoritmo consta de los siguientes pasos:

- Comenzando con m0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las N-m0 iteraciones, agregue un nodo nuevo que se conecte con m (< m0) enlaces a nodos preexistentes con probabilidad:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

la probabilidad de que un enlace del nuevo nodo se conecte al nodo i depende del grado ki como

Growth and Preferential Attachment



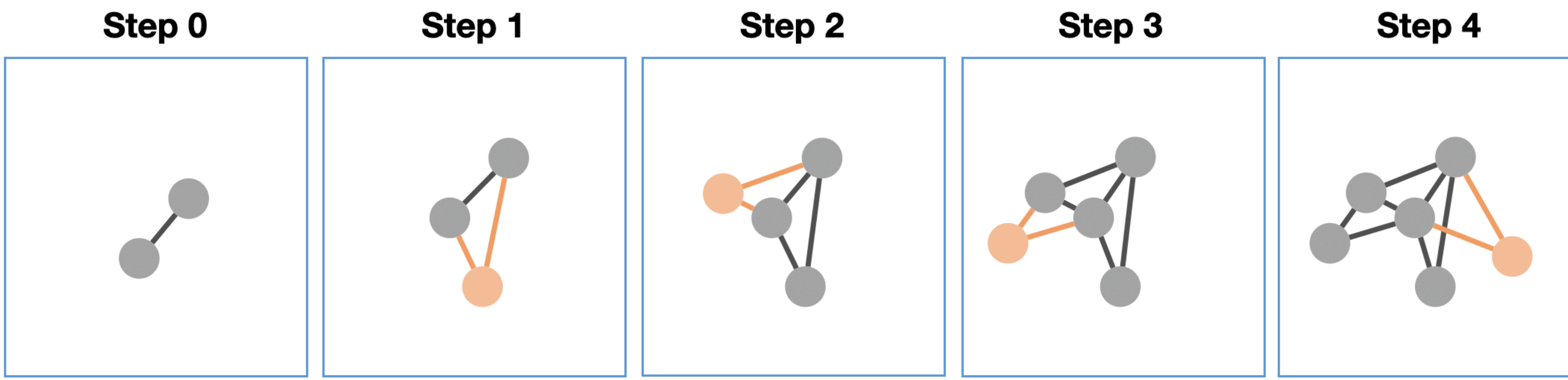
El algoritmo consta de los siguientes pasos:

- Comenzando con m0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las N-m0 iteraciones, agregue un nodo nuevo que se conecte con m (< m0) enlaces a nodos preexistentes con probabilidad:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

la probabilidad de que un enlace del nuevo nodo se conecte al nodo i depende del grado ki como

Growth and Preferential Attachment



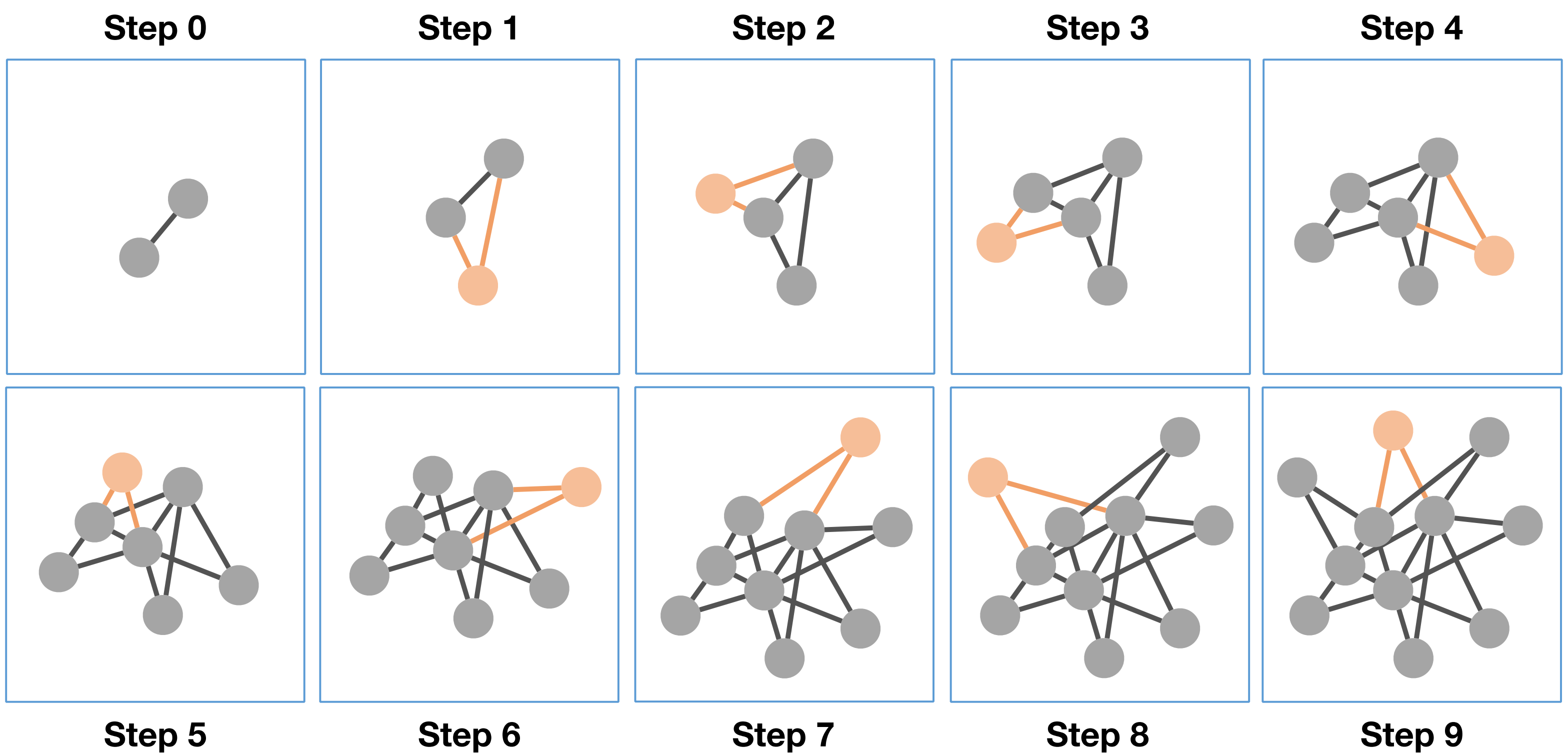
El algoritmo consta de los siguientes pasos:

- Comenzando con m0 nodos, los enlaces se eligen arbitrariamente, siempre que cada nodo tenga al menos un enlace.
- Para las N-m0 iteraciones, agregue un nodo nuevo que se conecte con m (< m0) enlaces a nodos preexistentes con probabilidad:

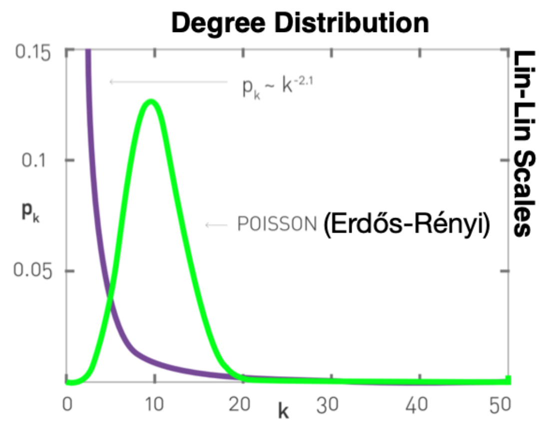
$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

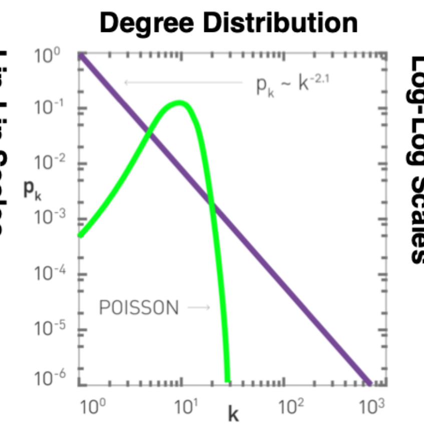
la probabilidad de que un enlace del nuevo nodo se conecte al nodo i depende del grado ki como

Growth and Preferential Attachment



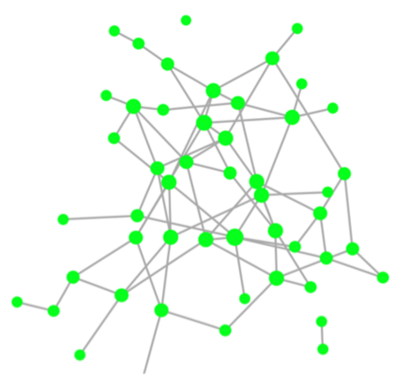
Scale-Free and Random Networks





¡Las redes sin escala tienen una distribución de grados que sigue una ley de potencia!

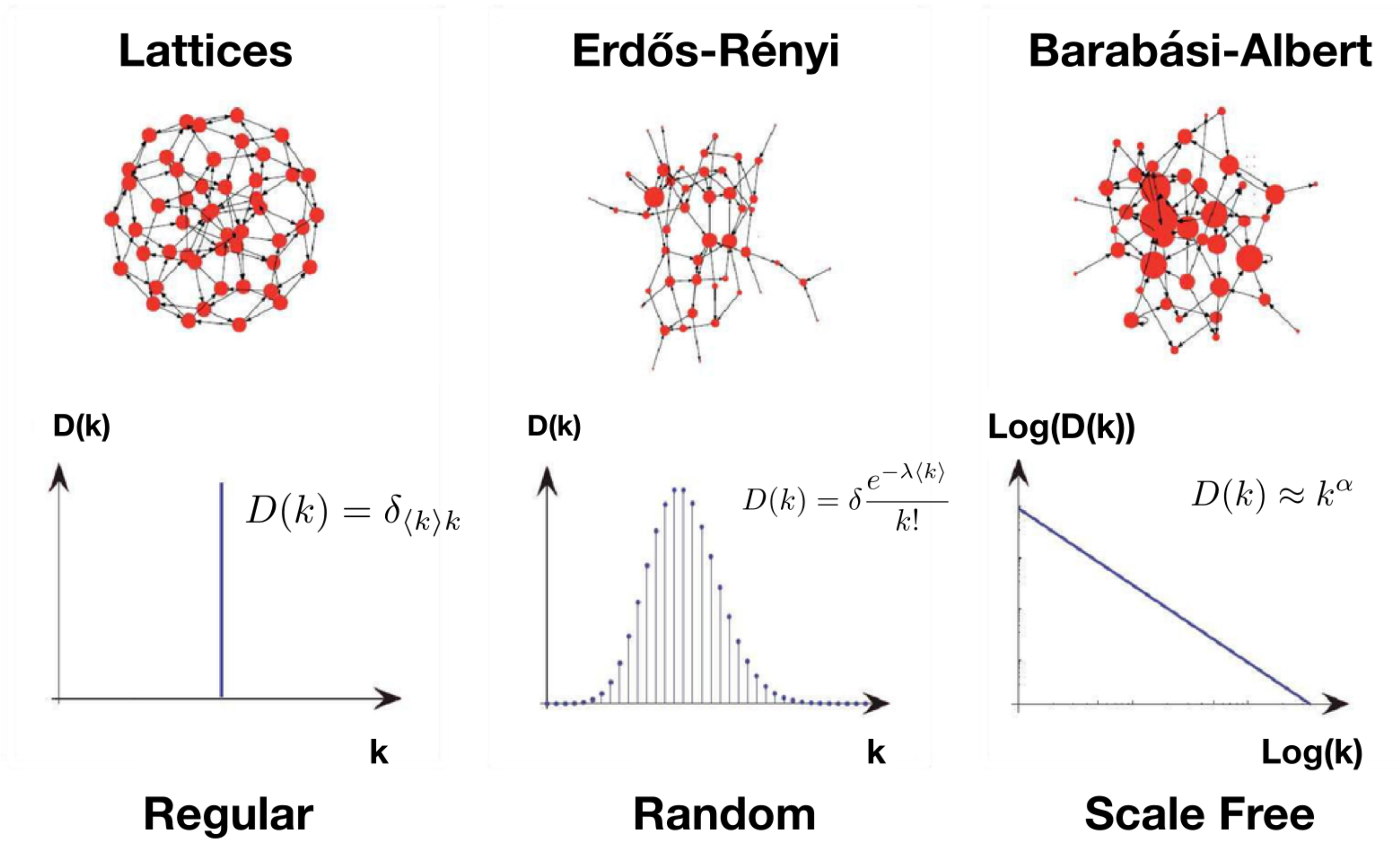
(c)



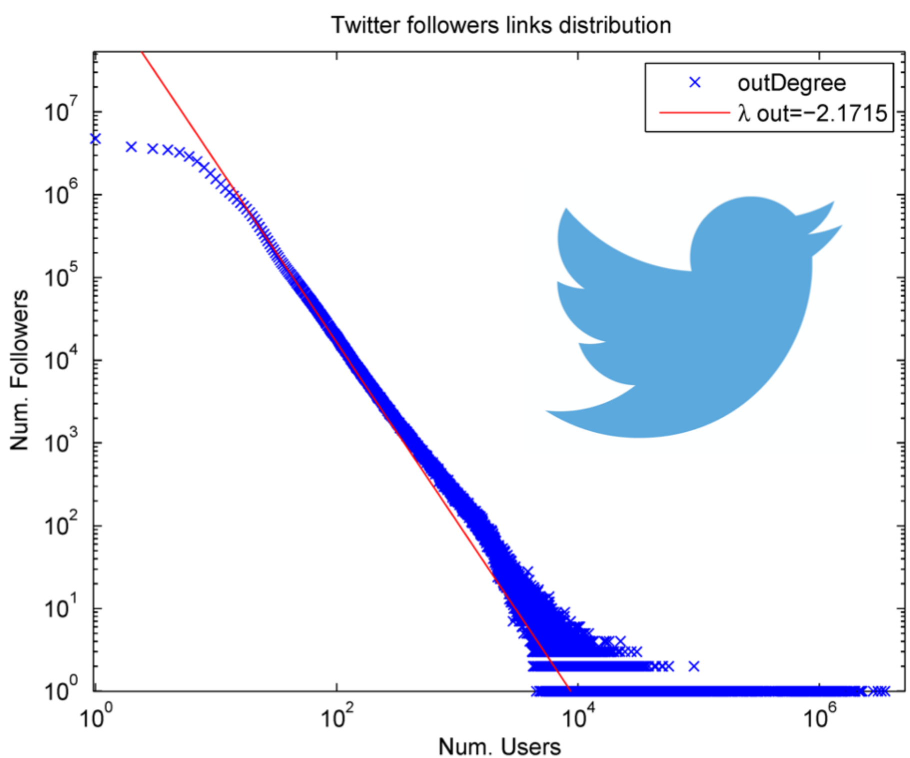
(Erdős-Rényi)

(d)

Scale-Free


Lattices
Erdős-Rényi
Barabási-Albert
D(k)
$D(k) = \delta_{\langle k \rangle k}$
k
D(k)
$D(k) = \delta \frac{e^{-\lambda \langle k \rangle}}{k!}$
k
Log(D(k))
$D(k) \approx k^{\alpha}$
Log(k)
Regular
Random
Scale Free

Scale Free Networks



- La ubicuidad de la propiedad sin escala sugiere algún mecanismo subyacente fundamental;

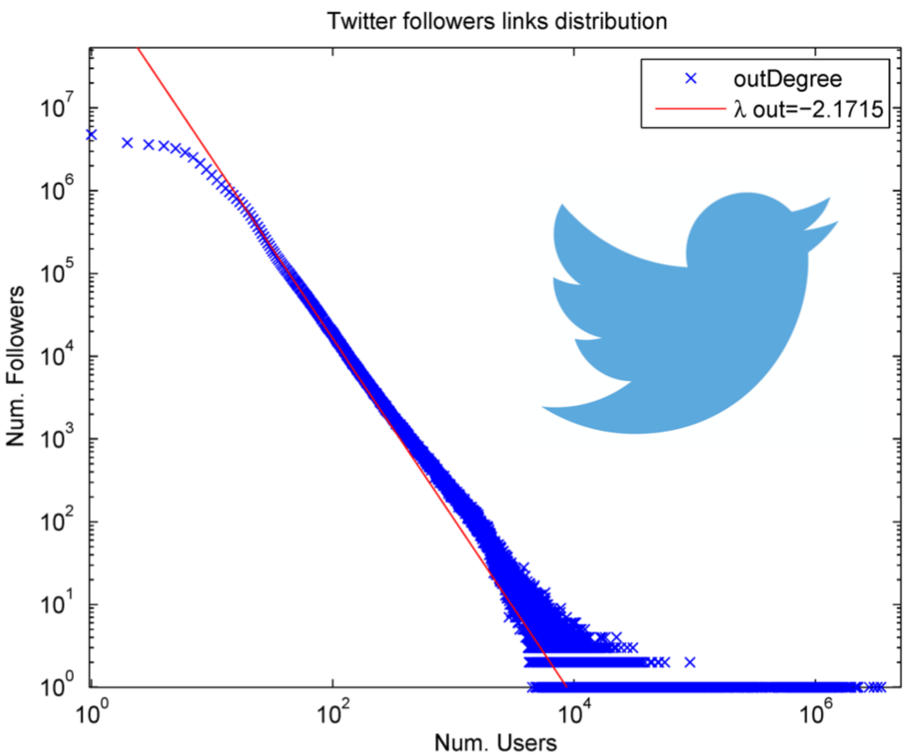
- Crecimiento y apego preferencial:
  - Crecimiento: se agregan nodos a la red con el tiempo;
  - Apego preferencial: estos nodos prefieren vincularse a nodos altamente conectados;

- El modelo simple de Barabási-Albert muestra que el crecimiento y el apego preferencial conducen a redes sin escala;

- Los ricos se hacen más ricos: los nodos altamente conectados tienden a seguir adquiriendo más conexiones

- Resultado claro del apego preferencial
  - En última instancia, conduce a una propiedad sin escala
  - Ventaja del primero: los nodos que se agregan antes tienen grados más altos

¿Es el modelo B.A. un buen modelo?

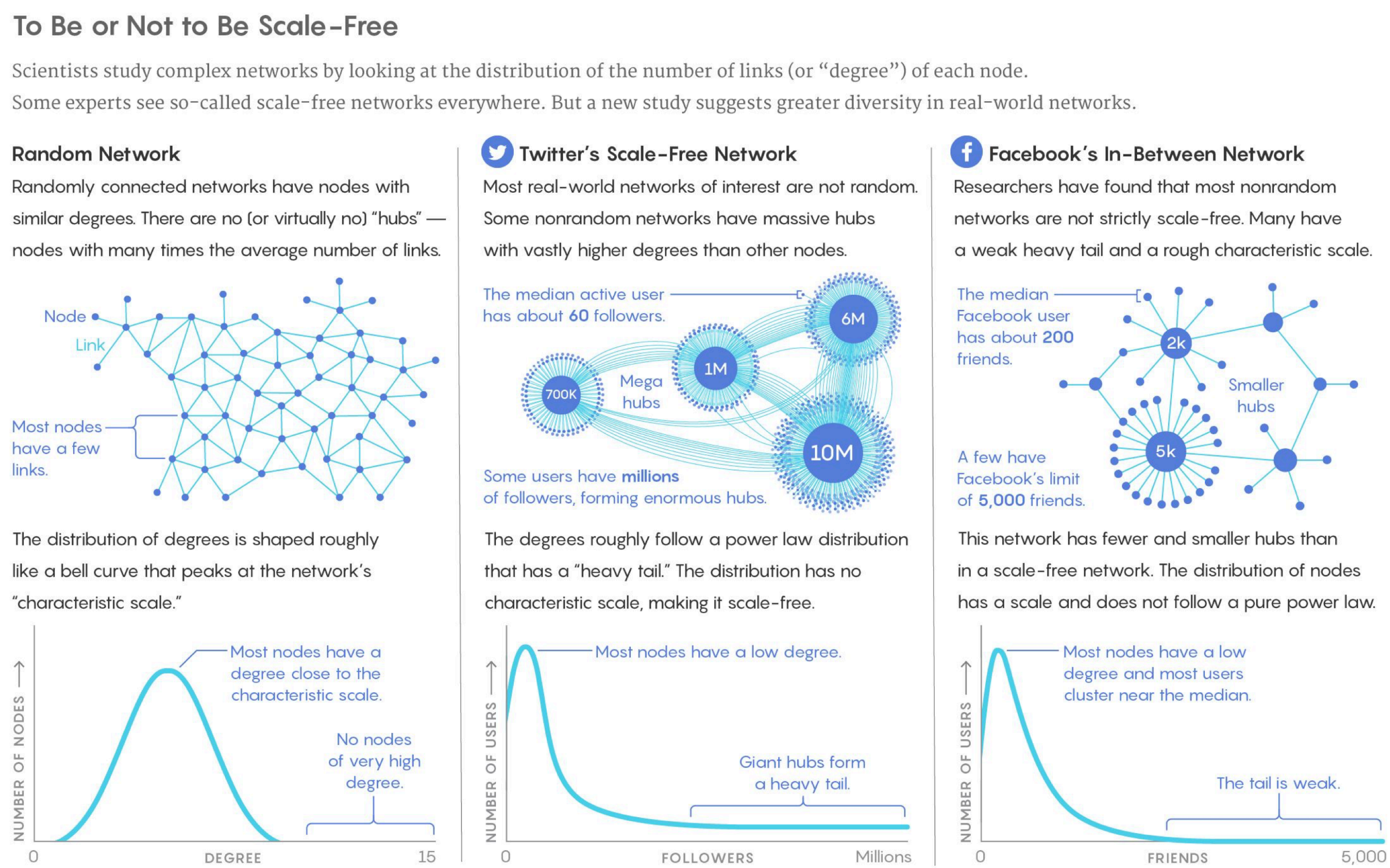


- Captura la intuición de que las redes reales crecen y que los nuevos nodos prefieren conectarse a nodos altamente conectados;
- Predice el comportamiento sin escala observado en redes reales;
- Que la preferencia sea directamente proporcional al número de enlaces parece simplista;
- La predicción de que los primeros nodos siempre están más conectados parece extraña;

Posibles extensiones:

- apego preferencial sublineal/superlineal
- métricas de fitness para nodos
- eliminación de nodos (edad fuera de la red)
- modelo mejor vinculado con la dinámica específica en cuestión.

“There is no general theory of networks”



https://www.quantamagazine.org/scant-evidence-of-power-laws-found-in-real-world-networks-20180215/?utm_content=buffer34d54&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer